

## Article

# Quantifying Next Generation Sequencing Sample Pre-Processing Bias in HIV-1 Complete Genome Sequencing

Bram Vrancken <sup>1,\*</sup>, Nídia Sequeira Trovão <sup>1</sup>, Guy Baele <sup>1</sup>, Eric van Wijngaerden <sup>2</sup>, Anne-Mieke Vandamme <sup>1,3</sup>, Kristel van Laethem <sup>1</sup> and Philippe Lemey <sup>1</sup>

Received: 28 September 2015; Accepted: 15 December 2015; Published: 7 January 2016

Academic Editors: Johnson Mak, Peter Walker and Marcus Thomas Gilbert

<sup>1</sup> Rega Institute for Medical Research, Clinical and Epidemiological Virology, Department of Microbiology and Immunology, KU Leuven—University of Leuven, 3000 Leuven, Belgium; Nidia.SequeiraTrovao@rega.kuleuven.be (N.S.T.); guy.baele@rega.kuleuven.be (G.B.); annemie.vandamme@uzleuven.be (A.-M.V.); kristel.vanlaethem@uzleuven.be (K.L.); philippe.lemey@rega.kuleuven.be (P.L.)

<sup>2</sup> University Hospitals Leuven, 3000 Leuven, Belgium; eric.vanwijngaerden@uzleuven.be

<sup>3</sup> Center for Global Health and Tropical Medicine, Unidade de Microbiologia, Instituto de Higiene e Medicina Tropical, Universidade Nova de Lisboa, 1349-008 Lisbon, Portugal

\* Correspondence: bram.vrancken@rega.kuleuven.be; Tel.: +32-16-33-28-92; Fax: +32-16-33-21-31

**Abstract:** Genetic analyses play a central role in infectious disease research. Massively parallelized “mechanical cloning” and sequencing technologies were quickly adopted by HIV researchers in order to broaden the understanding of the clinical importance of minor drug-resistant variants. These efforts have, however, remained largely limited to small genomic regions. The growing need to monitor multiple genome regions for drug resistance testing, as well as the obvious benefit for studying evolutionary and epidemic processes makes complete genome sequencing an important goal in viral research. In addition, a major drawback for NGS applications to RNA viruses is the need for large quantities of input DNA. Here, we use a generic overlapping amplicon-based near full-genome amplification protocol to compare low-input enzymatic fragmentation (Nextera™) with conventional mechanical shearing for Roche 454 sequencing. We find that the fragmentation method has only a modest impact on the characterization of the population composition and that for reliable results, the variation introduced at all steps of the procedure—from nucleic acid extraction to sequencing—should be taken into account, a finding that is also relevant for NGS technologies that are now more commonly used. Furthermore, by applying our protocol to deep sequence a number of pre-therapy plasma and PBMC samples, we illustrate the potential benefits of a near complete genome sequencing approach in routine genotyping.

**Keywords:** HIV; full genome sequencing; NGS

## 1. Introduction

Determining the most efficient antiretroviral treatment options through genotypic testing for the presence of resistance-associated mutations (RAMs) has become common practice in the routine management of HIV infection. Studies addressing the sensitivity and accuracy of standard genotypic tests have revealed that RAMs present at levels below 20% are generally not detected and that there is considerable uncertainty about the estimated proportion of detected RAMs that are present in less than 50% of the population [1–4]. This is a major limitation, because minority drug-resistant subpopulations can limit therapy efficacy [5–13]. In addition, the outgrowth of (undetected) minority

variants with RAMs under drug selective pressure increases the risk of transmission of drug resistance, which is also associated with therapy failure [14,15].

The integrated state of HIV is an important feature of the viral life cycle in the context of resistance to antiretroviral therapy, because it allows drug-resistant variants to persist as archived provirus in peripheral blood mononuclear cells (PBMCs). Consequently, even though drug resistance can be undetectable in the plasma compartment, this “memory” reservoir can exert a long-term impact on responses to antiretroviral therapy [16,17]. Yet, routine genotyping tests generally focus on currently circulating viral variants by examining the plasma fraction of blood.

Currently-used clinical drug resistance tests also focus on a limited number of small genomic fragments. There is, however, a growing need to simultaneously probe multiple genomic regions due to the development of drugs that target different steps in the viral life cycle [18]. In recognition of this need, studies that combine both the full genome and deep sequencing dimensions have already been reported [19–26]. Yet, the need for relatively large quantities of input DNA represents an important drawback for deep sequencing applications to RNA viruses. To meet these requirements, most researchers initially resorted to a sequence-specific amplification and sequencing of small genomic fragments following RT-PCR amplification of the target(s) of interest [27]. While this method allows the incorporation of platform-specific adaptors and, if required, barcodes for multiplexing during the PCR [28], it may introduce systematic biases, as well as stochastic error [29–32]. Small amplicon-based approaches cannot readily be extended to full genome sequencing, because targeting numerous overlapping small amplicons becomes impractically labor-intensive and cost prohibitive. This is further aggravated by the need to perform replicate analyses in order to minimize the potential impact of stochastic fluctuations [33,34]. Two alternatives have been suggested to at least partly address these problems. Willerth *et al.* [20] demonstrated that HIV genomes from cell culture supernatant can be amplified and sequenced on the Illumina platform using random-priming strategies, which were adopted from transcriptome analysis. Similar promising approaches directly target the viral RNA with improved versions of the same or other sequence-independent amplification protocols [25,26]. Whereas these approaches should be less prone to amplification-associated biases (but see [35]) and have proven useful, it remains to be established how well these systems perform on (challenging) clinical samples in a direct comparison to highly-optimized RT-PCR based methods, the current gold standard in routine clinical care. An alternative approach is to amplify only a small set of large overlapping amplicons covering the genome [19,21–23], albeit with potential biases due to sequence-specific amplification. Both sequence-specific and random amplification methods may generate fragments that are too long for efficient clonal amplification and sequencing. The sample DNA is typically fragmented to the correct size range using mechanical (sonication or nebulization) or enzymatic approaches (e.g., Nextera™ or NEBNext Fragmentase), but all of these suffer from inherent biases [36,37].

Here, we make use of a sequence-specific amplification strategy for in-depth complete genome characterization in clinical HIV-1 samples on the Roche 454 platform. We adopt a strategy similar to Bimber *et al.* [19] and Henn *et al.* [22] and generate six large overlapping amplicons to cover the entire HIV-1 genome, but similar to Gall *et al.* [21], we extend the applicability of this approach by implementing (RT-) PCR protocols optimized for the generic detection of HIV-1 group M (responsible for the worldwide epidemic) [38–42]. In the next step, we compare an enzymatic fragmentation method, the Nextera™ transposon-based technology that has low DNA input requirements, to conventional mechanical Roche 454 shearing. By examining pre-therapy plasma and PBMC samples from patients of a small transmission chain [43] and from a patient who failed first-line therapy because of an undetected low-level drug resistance mutation [5], we demonstrate the utility of our approach in hypothesis-driven research.

## 2. Experimental Section

### 2.1. Samples

We studied the viral population in a small HIV-1 subtype B transmission chain (involving three patients: AR01, AR05 and AR07) [43] and in one additional patient [5] based on three plasma samples, two PBMC pellets and the plasma-derived outer PCR product of amplicons gag-PR, p2-RNaseH, IN-Vif, Env and Nef. An overview of the available samples is presented in Table 1. The patients of the transmission chain were all infected by dual-class resistant HIV-1. Patient AR06 failed first line therapy because of an undetected minority variant. The research was conducted according to the Declaration of Helsinki, and the use of patient samples was approved by the medical ethics committee of the University Hospitals Leuven, Leuven, Belgium (Reference B322201420270).

**Table 1.** Overview of the available samples.

Patient	Plasma	PBMC Pellet	Outer PCR Product *
AR01	no	yes	yes
AR05	yes	no	yes
AR06	yes	no	no
AR07	yes	yes	yes

\* No outer PCR product for amplicon Vif-Vpr-Vpu was available for any of the samples (see also Supplementary Materials, Table S1).

### 2.2. RNA Extraction

A starting volume of 140 µL plasma was eluted in 60 µL of elution buffer from the QIAamp Viral RNA Mini Kit (Qiagen, Venlo, The Netherlands) as described in the manufacturer's spin protocol. Viral RNA was extracted in six replicates and pooled to minimize sampling effects [33] and stored at −80 °C.

### 2.3. DNA Extraction

DNA was extracted once using the QIAamp Blood DNA Mini Kit (Qiagen) according to the spin protocol. To ensure purification of RNA-free DNA, our protocol included the optional addition of RNase A (Qiagen).

### 2.4. Reverse Transcription and PCR Amplification

The coding part of the HIV-1 genome of the viral and proviral samples was amplified using a set of 6 overlapping amplicons (Supplementary Materials, Figure S1). (RT-) PCR amplifications were also replicated and pooled before proceeding to the next step (Supplementary Materials, Table S1) [34]. The strategy of pooling replicate RNA extractions ensured that we obtained sufficient volume for the 5-replicate cDNA synthesis of all amplicons, except for patient AR01. Due to the absence of stored plasma for this patient, we had to resort to a single plasma-derived outer PCR product (Table 1 and Supplementary Materials, Table S1).

The first round amplification of the PBMC samples was performed with the same master mixes as used for reverse transcription and amplification of plasma samples. The cycling conditions were also identical, except for the omission of the RT step (resulting in a hot-started cycling program). The extraction volume we obtained did not permit 5-replicate cDNA synthesis with the usual 10 µL extract as the input. Instead, to arrive at the same number of replicates, 6 µL of DNA extract were complemented with 4 µL of H<sub>2</sub>O, except for Nef. Here, we added 10 µL of the DNA extract, because the test with the 6-µL approach yielded no positive result as determined on a 1% agarose gel. Because of this, only 4 replicate outer PCR reactions were possible for Nef.

All reactions were performed in a Biometra T3000 thermal cycler. An overview of all (RT-)PCR mixes, cycling conditions and a list of primers (synthesized by Eurogentec) is provided in the Supplementary Materials, Table S2 and Figures S2 and S3. The quality of the PCR products was visually assessed through gel electrophoresis on a 1% agarose gel.

### 2.5. Purification and Quantification

PCR products from the plasma and PBMC samples were pooled, purified and quantified before further processing using the different fragmentation methods. The sample fraction for fragmentation using the general 454 library preparation was purified with the Illustra GFX PCR DNA and Gel Band Purification kit (GE Healthcare, Diegem, Belgium). The fraction fragmented with the Nextera<sup>TM</sup> protocol was purified with DNA Clean & Concentrator (Zymo Research, Freiburg, Germany). All quantifications were done with the Quant-iT dsDNA HS Assay Kit or Quant-iT dsDNA BR Assay Kit from Invitrogen (Waltham, MA, USA). Upon quantification, all amplicons from the same sample were pooled in an equimolar fashion.

### 2.6. Fragmentation

Five micrograms of the pooled inner PCR product of both the RNA and DNA samples was subjected to fragmentation by mechanical shearing according to the general 454 library preparation protocol (Roche Diagnostics, Vilvoorde, Belgium) while incorporating multiplex identifier tags (MIDs). The fragmentation reaction with the Nextera<sup>TM</sup> DNA Sample Prep kit (Roche Titanium compatible, Epicentre Biotechnologies, Madison, WI, USA) was performed according to the manufacturer's protocol and resulted in a sequencing-ready bar-coded library. The modest input requirement of 50 ng allowed the use of both the outer and inner PCR products of all samples for comparison, except for the plasma sample from patient AR01, because the stored outer PCR product did not contain sufficient DNA.

### 2.7. Sequencing and Data Analysis

Before proceeding with the sequencing, we assessed the quality of the libraries using capillary electrophoresis (Agilent BioAnalyzer, Agilent, Diegem, Belgium). Because the fragment size distribution of the Nextera<sup>TM</sup> libraries was skewed, we tested two emulsion PCR (emPCR) conditions (0.15 and 0.30 copies per bead (cpb)) on the inner PCR product of the PBMC sample from patient AR01. Due to a technical error, no deep sequencing data could be generated for the Nextera<sup>TM</sup> fragmented plasma inner PCR product of patient AR07. An overview of the number of reads for each sample is provided in the Supplementary Materials, Table S3.

Sequencing was carried out by Genomics Core (University Hospitals Leuven, Leuven, Belgium) on the GS-FLX 454 pyrosequencing platform (Roche Applied Science, Vilvoorde, Belgium) with Titanium chemistry, and the results were provided as Standard Flowgram Files. Sequence data were extracted and converted to FASTA and QUAL-format with a freely-available Python script [44]. During this format conversion, reads were clipped at the transition of high to low quality base calls at the site recommended by the 454 software. Prior to the read cleaning with RC454 [22], reads with an exact match to both the barcode and, for the Nextera<sup>TM</sup>-fragmented samples, the transposon end sequences were extracted with Segminator II [45] and assigned to the corresponding sample. When available, we used patient-specific reference sequences obtained by Sanger sequencing to map the reads (see Supplementary Materials, Table S4). *De novo* assembled sequences obtained by VICUNA [46] were used to map the reads in the remaining genome regions. We used the V-Phaser algorithm [22] in an attempt to distinguish sequencing errors from true variation.

## 3. Results

We first report on the degree of variability associated with the emPCR and sequencing steps. Next, we compare the variation between the Nextera<sup>TM</sup> fragmentation method and conventional

mechanical shearing in the sample comparisons. Finally, we construct a near complete genome resistance profile for clinical plasma and PBMC samples.

### 3.1. emPCR/Sequencing Associated Variability

Because of the cautionary approach to test two emPCR conditions and due to low coverages after the first run, a number of Nextera™ fragmented samples were clonally amplified and sequenced in duplicate (Supplementary Materials, Table S3). To score the concordance between the results from these duplicates, we calculated the fraction of positions at which the difference in detected frequency of all nucleotides is at most 1%, 5% and 10% (Table 2). These fractions are on average 87.94%, 98.73% and 99.72%, respectively. To visualize this variation, we plotted the largest difference in observed nucleotide frequency along the axis of the patient-specific reference sequence for these samples (Figures S4–S9). In accordance with the highly similar results, the majority rule consensus sequence differed at only 15 positions in the six samples under comparison (median: three; range: 1–4). In 12/15 (80%), this could be attributed to a nearly 50%-50% mixture of two variants, where a small difference can tip the balance in favor of one nucleotide. We also noted a few outliers where the difference in the detected proportion of a nucleotide amounts to  $\geq 20\%$ , which represent 0.01%–0.07% of all positions in the compared samples. Of these, eight (34.78%) are located within or adjacent ( $\pm 5$  nt) to homopolymers (length  $\geq 4$ ).

**Table 2.** Proportions of sites with nucleotide differences below 1%, 5% and 10% for various sample comparisons.

Patient	Sample	1%	5%	10%
emPCR/sequencing				
AR01	PBMC inner PCR <sup>a</sup>	86.78	99.41	99.92
AR05	plasma outer PCR	92.56	99.30	99.88
AR06	plasma outer PCR	85.69	98.72	99.84
	plasma inner PCR 1a-2a <sup>b</sup>	84.03	98.17	99.59
	plasma inner PCR 1b-2b <sup>b</sup>	87.89	98.38	99.63
AR07	PBMC outer PCR	88.21	98.61	99.65
	plasma outer PCR	90.48	98.51	99.53
	average	87.94	98.73	99.72
fragmentation method				
AR01	PBMC inner PCR	78.90	96.34	99.29
	plasma inner PCR	90.30	98.68	99.61
AR05	plasma inner PCR	88.90	99.14	99.80
AR06	plasma inner PCR	81.80	97.97	99.49
AR07	PBMC inner PCR	85.66	97.33	99.07
	average	85.11	97.89	99.45

Only positions with coverage  $\geq 100$  for both samples are taken into account. When possible, replicate sequence data were pooled when comparing the fragmentation methods. <sup>a</sup> Data for this sample were obtained from the same run, but with two emulsion PCR (emPCR) conditions (0.15 and 0.30 copies per bead (cpb)). <sup>b</sup> Due to a technical error, the inner PCR product of the plasma sample of patient AR06 was sequenced twice at both sequencing runs (with different multiplex identifier tags (MIDs)) (see also the Supplementary Materials, Table S3).

This variability likely stems from the random disproportional attachment of templates to empty beads during the emPCR step [47] or from errors arising during the actual sequencing process. Because many of the factors that determine the pyrosequencing error rate (e.g., position in the sequence, size of the template and spatial localization on the picotiter plate (PTP) [48]) vary between sequencing experiments, we considered this source of error as essentially stochastic. For this reason,



and similar to the strategy of pooling extracts and (RT-) PCR products, we pooled the sequence data for samples from both runs for the remaining analyses.

### 3.2. Comparison of Nextera™ with Standard Shearing

We compared the compositional differences between both fragmentation methods, Nextera™ and standard shearing, in the same way as above (Table 2). The fraction of sites where the largest difference in the detected percentage of any of the nucleotides amounts to 1%, 5% and 10% is on average 85.11%, 97.89% and 99.45%, respectively. Per category, this is 2.83%, 0.84% and 0.27% less when compared to the mean emPCR/sequencing-associated variability. The consensus sequences (majority rule) of the five Nextera™ and standard shearing fragmented samples differs at 20 positions (median: three; range: 1–9), which can be attributed to a nearly 50%-50% mixture of two nucleotides in 17 cases (85%). Positions with a difference in frequency of a nucleotide  $\geq 20$  represent 0.04% to 0.15% of all positions in the compared samples; 10 (50%) of these are associated with homopolymers (length  $\geq 4$ ). A visual representation of the variability of when comparing fragmentation methods is provided in the Supplementary Materials, Figures S10–S14.

### 3.3. Resistance Profiling

We investigated the potential of our near full genome genotyping approach by creating a RAM profile for each patient using the Stanford list with major HIV-1 drug resistance mutations [49], complemented with the list of reverse transcriptase (RT) and protease (PR) mutations used for drug resistance surveillance [50]. While the former list includes RAMs spread over the entire genome, the latter focuses on the protease and RT region and includes additional sentinel mutations. To illustrate that no (low-frequency) RAMs were detected at most positions associated with drug resistance, we provide the complete RAM profiles of the four patients in the Supplementary Materials, Tables S5–S8.

Data obtained by traditional cloning and Sanger sequencing were available for the PR and RT regions for all patients from the same sample that was used for 454 data generation, as was pre-therapy follow-up data for the transmission chain patients (Supplementary Materials, Tables S4–S7) [43]. A comparison highlights that all ubiquitously-present ( $\geq 80\%$ ) RAMs were detected in similar proportions irrespective of the sequencing and fragmentation technique. When contrasting the cloned and NGS data, most differences remain limited to 1%, and 10/16 (62.5%) were completely fixed according to all approaches. There are two notable outliers: nucleoside reverse transcriptase inhibitor (NRTI) position 41L and non-nucleoside reverse transcriptase inhibitor (NNRTI) position 179D of patient AR06, where the difference amounts to 4% and 10.28%, respectively. Some minority variant RAMs ( $\leq 20\%$ ) were only detected by the pyrosequencing approach ( $n = 12$ ; median: four; range: 2–6). In contrast, in each patient, one minority variant RAM was only picked up by the cloned sequence data. The resistance profiles obtained by Nextera™ fragmentation and standard shearing also are very similar, and all unique RAM calls again only involve RAMs at levels  $\leq 20\%$ .

The low input requirements of the Nextera™ fragmentation method enabled profiling the outer PCR products of a number of samples (Table 2 and Supplementary Materials, Table S3). When contrasted with the resistance profiles obtained from the matching Nextera™ fragmented inner PCR products, all differences in proportions of RAMs were below 5%.

To compare the diversity of the PBMC and plasma compartments, the amino acid content profiles following Nextera™ and standard shearing fragmentation were combined, and a histogram of the diversity per position was created (Supplementary Materials, Figures S15 and S16). This reveals a number of subpopulations in the PBMC compartment of patient AR01 that were not detected in the plasma sample. The higher diversity of the PBMC reservoir is reflected in a mean of 2.01 different amino acids per position, *vs.* 1.37 in the plasma. For patient AR07, the diversity of the PBMC reservoir is also higher, but less distinct from that of the plasma viruses: the average number of amino acids per position is 1.75 for the PBMC and 1.56 for the plasma reservoir. The resistance profiles for both

reservoirs reveal the same pattern for minority variants as observed in the previous comparisons. However, a number of moderately prevalent RAMs ( $\geq 20\%$ ) are only seen in the PBMC reservoir. Specifically, this concerns NRTI substitution 184I and fusion inhibitor substitution 36S for patient AR01, and for patient AR07, this involves NNRTI substitution 190E.

#### 4. Discussion

Massively parallel sequencing methods enable studying the composition of complex viral populations in a cost-effective fashion and point to the potential for near full-length genotyping approaches in routine clinical practice [51–54]. In order to take up an important role in decision-making in patient care, NGS-based genotyping approaches need to demonstrate excellent sensitivity and accuracy. It is therefore no surprise that much effort has been invested to investigate biases at all stages from sample to sequence data. Here, we compared two fragmentation methods, Nextera™ fragmentation and standard shearing, for a near complete genome characterization based on a new sequence-specific amplification protocol.

In agreement with earlier findings that show similar results for different fragmentation methods [55], we find a similar population composition for both fragmentation procedures. Nevertheless, we recommend the Nextera™ fragmentation, because the lower DNA input requirements allow sequencing the product of a single PCR amplification, which avoids the risk of biases associated with an additional sequence-specific amplification of nested PCR protocols. Interestingly, the DNA input requirement for Nextera™ has further decreased from 50 ng since the start of this study to 1 ng at present, which enables examining more challenging samples. Recently, it also became possible to anticipate elongation and sequencing errors in the experimental sample pre-processing procedures. Some of these procedures rely on attaching unique identifiers to each DNA fragment, either by tagging the primer [56] or sequencing adaptors [57] with a unique barcode. Others procedures rely on the rolling circle amplification of the input DNA or RNA molecules [58,59]. They all result in sequence reads that can be grouped either by the shared unique barcode or by the physical linkage and allow for creating consensus sequences for all reads that represent the same original template. An additional advantage of early-stage tagging [56] over later-stage tagging [57] and rolling circle amplification [58,59] is that the former can also identify template resampling and can be used to reconstruct the original population structure with greater confidence.

Replicate sequencing of some samples reveals that most of the variability results from the emPCR/sequencing steps rather than from the fragmentation methods (Table 2). The introduction of substantial variability after the fragmentation step in the sample pre-processing procedures highlights that in order to arrive at high levels of confidence in the observed frequencies of (minority) variants, it is crucial to average out random errors by pooling replicates at every stage of the pre-processing protocol and not only at the extraction and amplification stages, the importance of which was previously shown [33,34]. Specific attention should be paid to this when accurate frequency estimates of minority variants are required, such as in establishing the clinically-significant cutoff values. Although Roche announced it will phase out the 454-pyrosequencing platform, we believe these results remain relevant for other NGS technologies because of the stochastic nature inherent to the clonal amplification and sequencing steps. As these steps are conceptually the same for the Ion Torrent platform, we specifically anticipate similar issues for this technology. Many more sequence reads per sample can be generated by Illumina platforms, which is currently the most widely-used NGS method. Because this decreases the potential bias of stochastic events, we anticipate that random effects that occur during the bridge amplification and sequencing steps will be less pronounced. The third generation of sequencing technologies (TGS, such as the already available PacBio SMRT and Oxford Nanopore minION systems) aim at directly determining the nucleotide order in the fragments of interest, making the massively parallel clonal amplification step redundant [60]. This considerably simplifies the sample preparation protocols, which reduces to an extraction and amplification step to arrive at sufficient amounts of input material. With an

appropriate experimental setup (e.g., pooling of replicate extraction and amplification products [33,34]), the risk of introducing artifacts can be greatly reduced. Nonetheless, it is well acknowledged that there always remains some unexplained variation (the so-called “batch effect” [61,62]), and this may become important when investigating the lower end of the frequency distribution of variants, independent of the NGS technology.

The TGS technologies also offer longer read lengths, which can in theory lead to complete variant sequencing without the need for assembly. However, the reduced sensitivity of long-range PCRs and the current error rates (~4% for minION and ~11% for single pass long reads in SMRT sequencing) currently prevent a sensitive screening of complex virus populations. On the other hand, the excellent base calling accuracy that can be achieved with the SMRT circular consensus sequencing method, at the expense of read length, calls for a coupling with sequence-specific and random priming strategies. The current throughput on the PacBio Sequel systems may however be too low to obtain sufficient coverage to compensate for the high levels of host contamination in the samples with the latter approach.

The relatively modest impact of the fragmentation and subsequent steps in the sample preparation protocol propagate into the RAM profiles. However, while the variability remains relatively modest for the majority variants, the difference in detected proportion can rise to several fold changes for minority variants. It can therefore be anticipated that the variance in observed proportions of majority variant RAMs introduced at the post-amplification stage will only rarely lead to a qualitative difference in resistance profile interpretation. In contrast, this source of variation will likely become important when determining the clinically-significant cutoffs.

The close agreement in estimated RAM frequencies when comparing matching outer and inner Nextera™ fragmented PCR products (differences are smaller than 5%) may be a consequence of the pooling strategy that levels out the effect of random errors. It is interesting to note that the 184I minority variant RAM reported by Van Laethem *et al.* [5] for patient AR06 was observed in the outer PCR product at the same level (2%), yet disappeared in the inner PCR product. On the other hand, the 184V variant that was linked to therapy failure and detected at 2% [5] could not be identified based on the outer nor inner PCR product. This highlights that caution is needed when scoring the presence/absence of variants at levels close to the detection limit. In addition to experimental variation (see below), a plausible cause for such discrepancies lies in the assumption of uniform error rates for homo- and hetero-polymer regions by the error correction algorithm we used [22], which undoubtedly accounts for a large fraction of the false negative and false positive minority variants.

The fact that the transmission chain patients are still successful on their first-line therapy, whereas patient AR06 experienced therapy failure due to undetected minority variant RAMs seems to add to the conflicting reports on the clinical relevance of minority resistance variants (e.g., [5–13,63–65]). However, the view has recently emerged that their effect is determined by an interplay between the particular resistance mutations and the specific regimen (low *vs.* high genetic barrier to resistance). In particular, because a single mutation can already lead to high levels of resistance against NNRTIs [66], minority NNRTI RAMs are associated with an increased risk of first-line therapy failure [13,14,67–69]. In line with this, patient AR06 started on a therapy with a low genetic barrier to resistance [5], in contrast to the NRTI + protease inhibitor (PI)/ritonavir based first-line regimen of the transmission chain patients. Of note, a thorough characterization of the virus population upon therapy failure may lead to the discovery of additional resistance mutations and can therefore offer potentially actionable information [70–73]. Because of differences in extraction volume for traditional cloning and sequencing (1 mL) and our NGS protocol ( $6 \times 140 \mu\text{L}$ ), we cannot directly compare the RAM profiles obtained by the different procedures. In particular, we cannot exclude that the minority variant RAMs detected exclusively by the traditional cloning approach are false negatives in the deep sequencing approach.

Our deep sequencing results confirm the earlier finding that all three patients from the small transmission chain (AR01, AR05 and AR07) share dominant (*i.e.*, 100%) dual-class RAMs [43].



Because it is known that drug-resistant variants tend to revert to wild-type upon transmission, we also investigated the available PBMC populations of patients AR01 and AR07 in an attempt to establish which viral reservoir may be best suited to detect (transmitted) drug resistance (TDR) at this stage. The detection of additional (N)NRTI mutations at levels above 20% (184I, patient AR01 and 190E, patient AR07) and a fusion inhibitor mutation (36S, patient AR01) unique to the PBMC compartment highlights the potential advantage of PBMCs for genotyping and also illustrates the potential benefit of a whole genome approach to genotyping. The long-term pre-therapy follow-up data show that none of the above three RAMs and none of the minority variant RAMs from either reservoir evolved to replace the wild-type variants. Of note, four out of five majority variant RAMs persisted in the almost six-year pre-therapy follow-up period for patient AR01, as was first noted by Van Laethem *et al.* [43]. The PBMC compartment also contained a more diverse population in patient AR01 and, to a lesser extent, in patient AR07 than the corresponding plasma compartment. Such differences between patients may reflect a difference between time of infection and time of sampling. In addition, both PBMC populations showed signs of APOBEC3-induced hypermutation as evidenced by multiple mutations of tryptophan to stop codons. Of note, the 184I, 190E and 36S substitutions we reported all involve a G to A transition and are therefore perhaps APOBEC3-induced rather than transmitted RAMs.

The current interpretation of genotypic resistance testing in routine clinical care, which only takes into account the presence/absence of RAMs, is highly efficient in selecting optimal drug combinations. Several lines of evidence, however, suggest that the selection process can be further optimized by including information on the broader genetic background. For example, Boltz *et al.* [63] show that low-frequency nevirapine (NVP, a NNRTI) RAMs that were selected for under single dose NVP drug selective pressure increase the risk of failing a subsequent NVP-based therapy, which was not observed for the same RAMs that did not emerge under NVP selection pressure. This can be explained by RAMs emerging upon single dose NVP exposure that are linked on the same viral genome. Similarly, coevolution of protease and its substrate Gag during protease inhibitor (PI) exposure can affect PI-based therapy [74,75]. Haplotype reconstruction of NGS data may help to address this, but this remains challenging for short read data [76]. An interesting avenue for further research may be the use of very long-read NGS technologies, such as the PacBio SMRT platform, to reconstruct the haplotypes (e.g., [77]), provided appropriate measures are taken to avoid both PCR-induced [78,79] and *in silico* recombination [76].

In summary, we find that variation introduced during the fragmentation and later steps of the sample preparation can impact the recovery of variant frequencies and that this mostly affects the prevalence estimates of minority variants. Furthermore, our generic near complete genome amplification approach may prove useful in large-scale phylodynamics studies and, after appropriate validation, perhaps also in further studies of the potential benefits of NGS in clinical decision making.

**Acknowledgments:** Bram Vrancken was supported by a PhD grant from the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen). Nidia Sequeira Trovão was supported by a grant from the Agência Nacional PROALV. This work was made possible by funding of the Onderzoeksfonds KU Leuven/Research Fund KU Leuven (Program Financing No. PF/10/018), by the Bijzonder Onderzoeksfonds KU Leuven (BOF) No. OT/14/115, by the Fonds voor Wetenschappelijk Onderzoek Vlaanderen (FWO) (Krediet Nos. 1.5.236.11N, 1.5.252.12N, G.0692.14N and G.0662.15N) and by the AIDS Reference Laboratory of Leuven that receives support from the Belgian Ministry of Social Affairs through a fund within the Health Insurance System. The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under Grant Agreement No. 278433-PREDEMICS and ERC Grant Agreement No. 260864. The VIROGENESIS project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 634650. We thank Y. Schrooten for outstanding technical assistance. We also thank two anonymous reviewers and the academic editor for insightful criticisms and comments.

**Author Contributions:** Bram Vrancken, Nidia Sequeira Trovão, Kristel van Laethem, Anne-Mieke Vandamme, Eric van Wijngaerden and Philippe Lemey conceived of the experiments. Bram Vrancken and Nidia Sequeira Trovão performed the wet-lab experiments. Bram Vrancken and Guy Baele performed the data analyses. All authors contributed to writing the manuscript.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Palmer, S.; Kearney, M.; Maldarelli, F.; Halvas, E.K.; Bixby, C.J.; Bazmi, H.; Rock, D.; Falloon, J.; Davey, R.T., Jr.; Dewar, R.L.; *et al.* Multiple, linked human immunodeficiency virus type 1 drug resistance mutations in treatment-experienced patients are missed by standard genotype analysis. *J. Clin. Microbiol.* **2005**, *43*, 406–413.
- Schuurman, R.; Demeter, L.; Reichelderfer, P.; Tijnagel, J.; de Groot, T.; Boucher, C. Worldwide evaluation of DNA sequencing approaches for identification of drug resistance mutations in the human immunodeficiency virus type 1 reverse transcriptase. *J. Clin. Microbiol.* **1999**, *37*, 2291–2296.
- Halvas, E.K.; Aldrovandi, G.M.; Balfe, P.; Beck, I.A.; Boltz, V.F.; Coffin, J.M.; Frenkel, L.M.; Hazelwood, J.D.; Johnson, V.A.; Kearney, M.; *et al.* Blinded, multicenter comparison of methods to detect a drug-resistant mutant of human immunodeficiency virus type 1 at low frequency. *J. Clin. Microbiol.* **2006**, *44*, 2612–2614.
- Schuurman, R.; Brambilla, D.; de Groot, T.; Huang, D.; Land, S.; Bremer, J.; Benders, I.; Boucher, C.A.B.; ENVA Working Group. Underestimation of HIV type 1 drug resistance mutations: Results from the ENVA-2 genotyping proficiency program. *AIDS Res. Hum. Retroviruses* **2002**, *18*, 243–248.
- Van Laethem, K.; de Munter, P.; Schrooten, Y.; Verbesselt, R.; van Ranst, M.; van Wijngaerden, E.; Vandamme, A.M. No response to first-line tenofovir+lamivudine+efavirenz despite optimization according to baseline resistance testing: Impact of resistant minority variants on efficacy of low genetic barrier drugs. *J. Clin. Virol.* **2007**, *39*, 43–47.
- Johnson, J.A.; Geretti, A.M. Low-frequency HIV-1 drug resistance mutations can be clinically significant but must be interpreted with caution. *J. Antimicrob. Chemother.* **2010**, *65*, 1322–1326.
- Johnson, J.A.; Li, J.F.; Wei, X.; Lipscomb, J.; Irlbeck, D.; Craig, C.; Smith, A.; Bennett, D.E.; Monsour, M.; Sandstrom, P.; *et al.* Minority HIV-1 drug resistance mutations are present in antiretroviral treatment-naïve populations and associate with reduced treatment efficacy. *PLoS Med.* **2008**, *5*, doi:10.1371/journal.pmed.0050158.
- Delobel, P.; Saliou, A.; Nicot, F.; Dubois, M.; Trancart, S.; Tangre, P.; Aboulker, J.P.; Taburet, A.M.; Molina, J.M.; Massip, P.; *et al.* Minor HIV-1 variants with the K103N resistance mutation during intermittent efavirenz-containing antiretroviral therapy and virological failure. *PLoS ONE* **2011**, *6*, doi:10.1371/journal.pone.0021655.
- Simen, B.B.; Simons, J.F.; Hullsiek, K.H.; Novak, R.M.; Macarthur, R.D.; Baxter, J.D.; Huang, C.; Lubeski, C.; Turenchalk, G.S.; Braverman, M.S.; *et al.* Low-abundance drug-resistant viral variants in chronically HIV-infected, antiretroviral treatment-naïve patients significantly impact treatment outcomes. *J. Infect. Dis.* **2009**, *199*, 693–701.
- Lataillade, M.; Chiarella, J.; Yang, R.; Schnittman, S.; Wirtz, V.; Mancini, M.; Uy, J.; Seekins, D.; Krystal, M.; McGrath, D.; *et al.* Prevalence and Clinical Significance of Transmitted Drug-Resistant (TDR) HIV Mutations by Ultra-Deep Sequencing (UDS) in HIV-Infected ARV-Naïve Subjects in CASTLE Study; Antiviral Therapy: London, UK, 2009; Volume 14, p. A44.
- Stekler, J.D.; Ellis, G.M.; Carlsson, J.; Eilers, B.; Holte, S.; Maenza, J.; Stevens, C.E.; Collier, A.C.; Frenkel, L.M. Prevalence and impact of minority variant drug resistance mutations in primary HIV-1 infection. *PLoS ONE* **2011**, *6*, doi:10.1371/journal.pone.0028952.
- Peuchant, O.; Thiébaud, R.; Capdepon, S.; Lavignolle-Aurillac, V.; Neau, D.; Morlat, P.; Dabis, F.; Fleury, H.; Masquelier, B.; ANRS CO3 Aquitaine Cohort. Transmission of HIV-1 minority-resistant variants and response to first-line antiretroviral therapy. *AIDS* **2008**, *22*, 1417–1423.
- Cozzi-Lepri, A.; Noguera-Julian, M.; di Giallonardo, F.; Schuurman, R.; Däumer, M.; Aitken, S.; Ceccherini-Silberstein, F.; D’Arminio Monforte, A.; Geretti, A.M.; Booth, C.L.; *et al.* Low-frequency drug-resistant HIV-1 and risk of virological failure to first-line NNRTI-based ART: A multicohort European case-control study using centralized ultrasensitive 454 pyrosequencing. *J. Antimicrob. Chemother.* **2015**, *70*, 930–940.

14. Wittkop, L.; Günthard, H.F.; de Wolf, F.; Dunn, D.; Cozzi-Lepri, A.; de Luca, A.; Kücherer, C.; Obel, N.; von Wyl, V.; Masquelier, B.; *et al.* Effect of transmitted drug resistance on virological and immunological response to initial combination antiretroviral therapy for HIV (EuroCoord-CHAIN joint project): A European multicohort study. *Lancet Infect. Dis.* **2011**, *11*, 363–371.
15. Little, S.J.; Holte, S.; Routy, J.P.; Daar, E.S.; Markowitz, M.; Collier, A.C.; Koup, R.A.; Mellors, J.W.; Connick, E.; Conway, B.; *et al.* Antiretroviral-drug resistance among patients recently infected with HIV. *N. Engl. J. Med.* **2002**, *347*, 385–394.
16. Booth, C.L.; Geretti, A.M. Prevalence and determinants of transmitted antiretroviral drug resistance in HIV-1 infection. *J. Antimicrob. Chemother.* **2007**, *59*, 1047–1056.
17. Gao, F.; Dongning, W. Minor-drug-resistant HIV populations and treatment failure. *Future Virol.* **2007**, *2*, 293–302.
18. Cane, P.A. New developments in HIV drug resistance. *J. Antimicrob. Chemother.* **2009**, *64*, i37–i40.
19. Bimber, B.N.; Dudley, D.M.; Lauck, M.; Becker, E.A.; Chin, E.N.; Lank, S.M.; Grunenwald, H.L.; Caruccio, N.C.; Maffitt, M.; Wilson, N.A.; *et al.* Whole-genome characterization of human and simian immunodeficiency virus intrahost diversity by ultradeep pyrosequencing. *J. Virol.* **2010**, *84*, 12087–12092.
20. Willerth, S.M.; Pedro, H.A.M.; Pachter, L.; Humeau, L.M.; Arkin, A.P.; Schaffer, D.V. Development of a low bias method for characterizing viral populations using next generation sequencing technology. *PLoS ONE* **2010**, *5*, doi:10.1371/journal.pone.0013564.
21. Gall, A.; Ferns, B.; Morris, C.; Watson, S.; Cotten, M.; Robinson, M.; Berry, N.; Pillay, D.; Kellam, P. Universal amplification, next-generation sequencing, and assembly of HIV-1 genomes. *J. Clin. Microbiol.* **2012**, *50*, 3838–3844.
22. Henn, M.R.; Boutwell, C.L.; Charlebois, P.; Lennon, N.J.; Power, K.A.; Macalalad, A.R.; Berlin, A.M.; Malboeuf, C.M.; Ryan, E.M.; Gnerre, S.; *et al.* Whole genome deep sequencing of HIV-1 reveals the impact of early minor variants upon immune recognition during acute infection. *PLoS Pathog.* **2012**, *8*, doi:10.1371/journal.ppat.1002529.
23. Gibson, R.M.; Meyer, A.M.; Winner, D.; Archer, J.; Feyertag, F.; Ruiz-Mateos, E.; Leal, M.; Robertson, D.L.; Schmotzer, C.L.; Quiñones-Mateu, M.E. Sensitive deep-sequencing-based HIV-1 genotyping assay to simultaneously determine susceptibility to protease, reverse transcriptase, integrase, and maturation inhibitors, as well as HIV-1 coreceptor tropism. *Antimicrob. Agents Chemother.* **2014**, *58*, 2167–2185.
24. Cuypers, L.; Snoeck, J.; Vrancken, B.; Kerremans, L.; Vuagniaux, G.; Verbeeck, J.; Nevens, F.; Camacho, R.J.; Vandamme, A.M.; van Dooren, S. A near-full length genotypic assay for HCV1b. *J. Virol. Methods* **2014**, *209*, 126–135.
25. Batty, E.M.; Wong, T.H.N.; Trebes, A.; Argoud, K.; Attar, M.; Buck, D.; Ip, C.L.C.; Golubchik, T.; Cule, M.; Bowden, R.; *et al.* A modified RNA-Seq approach for whole genome sequencing of RNA viruses from faecal and blood samples. *PLoS ONE* **2013**, *8*, doi:10.1371/journal.pone.0066129.
26. Malboeuf, C.M.; Yang, X.; Charlebois, P.; Qu, J.; Berlin, A.M.; Casali, M.; Pesko, K.N.; Boutwell, C.L.; DeVincenzo, J.P.; Ebel, G.D.; *et al.* Complete viral RNA genome sequencing of ultra-low copy samples by sequence-independent amplification. *Nucleic Acids Res.* **2013**, *41*, doi:10.1093/nar/gks794.
27. Vrancken, B.; Lequime, S.; Theys, K.; Lemey, P. Covering all bases in HIV research: Unveiling a hidden world of viral evolution. *AIDS Rev.* **2010**, *12*, 89–102.
28. Hoffmann, C.; Minkah, N.; Leipzig, J.; Wang, G.; Arens, M.Q.; Tebas, P.; Bushman, F.D. DNA bar coding and pyrosequencing to identify rare HIV drug resistance mutations. *Nucleic Acids Res.* **2007**, *35*, doi:10.1093/nar/gkm435.
29. Karrer, E.E.; Lincoln, J.E.; Hogenhout, S.; Bennett, A.B.; Bostock, R.M.; Martineau, B.; Lucas, W.J.; Gilchrist, D.G.; Alexander, D. *In situ* isolation of mRNA from individual plant cells: Creation of cell-specific cDNA libraries. *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 3814–3818.
30. Polz, M.F.; Cavanaugh, C.M. Bias in template-to-product ratios in multitemplate PCR. *Appl. Environ. Microbiol.* **1998**, *64*, 3724–3730.
31. Vrancken, B.; Lemey, P. High-throughput HIV sequencing: Evolution in 2D. *Future Virol.* **2011**, *6*, 417–420.
32. Bracho, M.A.; García-Robles, I.; Jiménez, N.; Torres-Puente, M.; Moya, A.; González-Candelas, F. Effect of oligonucleotide primers in determining viral variability within hosts. *Virol. J.* **2004**, *1*, doi:10.1186/1743-422X-1-13.

33. Poon, A.F.Y.; Swenson, L.C.; Dong, W.W.Y.; Deng, W.; Kosakovsky Pond, S.L.; Brumme, Z.L.; Mullins, J.I.; Richman, D.D.; Harrigan, P.R.; Frost, S.D.W. Phylogenetic analysis of population-based and deep sequencing data to identify coevolving sites in the nef gene of HIV-1. *Mol. Biol. Evol.* **2010**, *27*, 819–832.
34. Vandenbroucke, I.; Marck, H.V.; Mostmans, W.; Eygen, V.V.; Rondelez, E.; Thys, K.; van Baelen, K.; Fransen, K.; Vaira, D.; Kabeya, K.; *et al.* HIV-1 V3 envelope deep sequencing for clinical plasma specimens failing in phenotypic tropism assays. *AIDS Res. Ther.* **2010**, *7*, doi:10.1186/1742-6405-7-4.
35. Hansen, K.D.; Brenner, S.E.; Dudoit, S. Biases in Illumina transcriptome sequencing caused by random hexamer priming. *Nucleic Acids Res.* **2010**, *38*, doi:10.1093/nar/gkq224.
36. Poptsova, M.S.; Il'icheva, I.A.; Nechipurenko, D.Y.; Panchenko, L.A.; Khodikov, M.V.; Oparina, N.Y.; Polozov, R.V.; Nechipurenko, Y.D.; Grokhovsky, S.L. Non-random DNA fragmentation in next-generation sequencing. *Sci. Rep.* **2014**, *4*, doi:10.1038/srep04532.
37. Marine, R.; Polson, S.W.; Ravel, J.; Hatfull, G.; Russell, D.; Sullivan, M.; Syed, F.; Dumas, M.; Wommack, K.E. Evaluation of a transposase protocol for rapid generation of shotgun high-throughput sequencing libraries from nanogram quantities of DNA. *Appl. Environ. Microbiol.* **2011**, *77*, 8071–8079.
38. Covens, K.; Dekeersmaeker, N.; Schrooten, Y.; Weber, J.; Schols, D.; Quiñones-Mateu, M.E.; Vandamme, A.M.; van Laethem, K. Novel recombinant virus assay for measuring susceptibility of human immunodeficiency virus type 1 group M subtypes to clinically approved drugs. *J. Clin. Microbiol.* **2009**, *47*, 2232–2242.
39. Snoeck, J.; Riva, C.; Steegen, K.; Schrooten, Y.; Maes, B.; Vergne, L.; van Laethem, K.; Peeters, M.; Vandamme, A.M. Optimization of a genotypic assay applicable to all human immunodeficiency virus type 1 protease and reverse transcriptase subtypes. *J. Virol. Methods* **2005**, *128*, 47–53.
40. Van Laethem, K.; Schrooten, Y.; Covens, K.; Dekeersmaeker, N.; de Munter, P.; van Wijngaerden, E.; van Ranst, M.; Vandamme, A.M. A genotypic assay for the amplification and sequencing of integrase from diverse HIV-1 group M subtypes. *J. Virol. Methods* **2008**, *153*, 176–181.
41. Van Laethem, K.; Schrooten, Y.; Dedeker, S.; van Heeswijck, L.; Deforche, K.; van Wijngaerden, E.; van Ranst, M.; Vandamme, A.M. A genotypic assay for the amplification and sequencing of gag and protease from diverse human immunodeficiency virus type 1 group M subtypes. *J. Virol. Methods* **2006**, *132*, 181–186.
42. Van Laethem, K.; Schrooten, Y.; Vandamme, A.M. In-house developed amplification protocols for Vif-Vpr-Vpu and Nef. Unpublished data, 2015.
43. Van Laethem, K.; Schrooten, Y.; Lemey, P.; Covens, K.; Dekeersmaeker, N.; van Ranst, M.; van Wijngaerden, E.; Vandamme, A.M. Transmission cluster of dual-class resistant HIV-1 in untreated patients. In Proceedings of The 13th International Bioinformatics Workshop on Virus Evolution and Molecular Epidemiology, Lisbon, Portugal, 9–14 September 2007.
44. Bioinformatics at COMAV. Available online: <https://bioinf.comav.upv.es/> (accessed on 6 January 2016).
45. Archer, J.; Baillie, G.; Watson, S.J.; Kellam, P.; Rambaut, A.; Robertson, D.L. Analysis of high-depth sequence data for studying viral diversity: A comparison of next generation sequencing platforms using Segminator II. *BMC Bioinform.* **2012**, *13*, doi:10.1186/1471-2105-13-47.
46. Yang, X.; Charlebois, P.; Gnerre, S.; Coole, M.G.; Lennon, N.J.; Levin, J.Z.; Qu, J.; Ryan, E.M.; Zody, M.C.; Henn, M.R. *De novo* assembly of highly diverse viral populations. *BMC Genom.* **2012**, *13*, doi:10.1186/1471-2164-13-475.
47. Gomez-Alvarez, V.; Teal, T.K.; Schmidt, T.M. Systematic artifacts in metagenomes from complex microbial communities. *ISME J.* **2009**, *3*, 1314–1317.
48. Gilles, A.; Megléc, E.; Pech, N.; Ferreira, S.; Malausa, T.; Martin, J.F. Accuracy and quality assessment of 454 GS-FLX Titanium pyrosequencing. *BMC Genom.* **2011**, *12*, doi:10.1186/1471-2164-12-245.
49. Shafer, R.W. Rationale and uses of a public HIV drug-resistance database. *J. Infect. Dis.* **2006**, *194*, S51–S58.
50. Bennett, D.E.; Camacho, R.J.; Otelea, D.; Kuritzkes, D.R.; Fleury, H.; Kiuchi, M.; Heneine, W.; Kantor, R.; Jordan, M.R.; Schapiro, J.M.; *et al.* Drug resistance mutations for surveillance of transmitted HIV-1 drug-resistance: 2009 Update. *PLoS ONE* **2009**, *4*, doi:10.1371/journal.pone.0004724.
51. Bellocave, P.; Recordon-Pinson, P.; Papuchon, J.; Vandenhende, M.A.; Reigadas, S.; Tauzin, B.; Fleury, H. Detection of low-frequency HIV type 1 reverse transcriptase drug resistance mutations by ultradeep sequencing in naive HIV type 1-infected individuals. *AIDS Res. Hum. Retroviruses* **2014**, *30*, 170–173.

52. Garcia-Diaz, A.; Guerrero-Ramos, A.; McCormick, A.L.; Macartney, M.; Conibear, T.; Johnson, M.A.; Haque, T.; Webster, D.P. Evaluation of the Roche prototype 454 HIV-1 ultradeep sequencing drug resistance assay in a routine diagnostic laboratory. *J. Clin. Virol.* **2013**, *58*, 468–473.
53. Quiñones-Mateu, M.E.; Avila, S.; Reyes-Teran, G.; Martinez, M.A. Deep sequencing: Becoming a critical tool in clinical virology. *J. Clin. Virol.* **2014**, *61*, 9–19.
54. Van Laethem, K.; Theys, K.; Vandamme, A.M. HIV-1 genotypic drug resistance testing: Digging deep, reaching wide? *Curr. Opin. Virol.* **2015**, *14*, 16–23.
55. Knierim, E.; Lucke, B.; Schwarz, J.M.; Schuelke, M.; Seelow, D. Systematic comparison of three methods for fragmentation of long-range PCR products for next generation sequencing. *PLoS ONE* **2011**, *6*, doi:10.1371/journal.pone.0028240.
56. Jabara, C.B.; Jones, C.D.; Roach, J.; Anderson, J.A.; Swanstrom, R. Accurate sampling and deep sequencing of the HIV-1 protease gene using a Primer ID. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 20166–20171.
57. Schmitt, M.W.; Kennedy, S.R.; Salk, J.J.; Fox, E.J.; Hiatt, J.B.; Loeb, L.A. Detection of ultra-rare mutations by next-generation sequencing. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 14508–14513.
58. Lou, D.I.; Hussmann, J.A.; McBee, R.M.; Acevedo, A.; Andino, R.; Press, W.H.; Sawyer, S.L. High-throughput DNA sequencing errors are reduced by orders of magnitude using circle sequencing. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 19872–19877.
59. Acevedo, A.; Andino, R. Library preparation for highly accurate population sequencing of RNA viruses. *Nat. Protoc.* **2014**, *9*, 1760–1769.
60. Morey, M.; Fernández-Marmiesse, A.; Castiñeiras, D.; Fraga, J.M.; Couce, M.L.; Cocho, J.A. A glimpse into past, present, and future DNA sequencing. *Mol. Genet. Metab.* **2013**, *110*, 3–24.
61. Leek, J.T.; Scharpf, R.B.; Bravo, H.C.; Simcha, D.; Langmead, B.; Johnson, W.E.; Geman, D.; Baggerly, K.; Irizarry, R.A. Tackling the widespread and critical impact of batch effects in high-throughput data. *Nat. Rev. Genet.* **2010**, *11*, 733–739.
62. Robasky, K.; Lewis, N.E.; Church, G.M. The role of replicates for error mitigation in next-generation sequencing. *Nat. Rev. Genet.* **2014**, *15*, 56–62.
63. Boltz, V.F.; Bao, Y.; Lockman, S.; Halvas, E.K.; Kearney, M.F.; McIntyre, J.A.; Schooley, R.T.; Hughes, M.D.; Coffin, J.M.; Mellors, J.W.; *et al.* Low-frequency nevirapine (NVP)-resistant HIV-1 variants are not associated with failure of antiretroviral therapy in women without prior exposure to single-dose NVP. *J. Infect. Dis.* **2014**, *209*, 703–710.
64. Balduin, M.; Oette, M.; Däumer, M.P.; Hoffmann, D.; Pfister, H.J.; Kaiser, R. Prevalence of minor variants of HIV strains at reverse transcriptase position 103 in therapy-naïve patients and their impact on the virological failure. *J. Clin. Virol.* **2009**, *45*, 34–38.
65. Vingerhoets, J.; Rimsky, L.; van Eygen, V.; Nijs, S.; Vanveggel, S.; Boven, K.; Picchio, G. Pre-existing mutations in the rilpivirine Phase III trials ECHO and THRIVE: Prevalence and impact on virological response. *Antivir. Ther.* **2013**, *18*, 253–256.
66. Wensing, A.M.; Calvez, V.; Günthard, H.F.; Johnson, V.A.; Paredes, R.; Pillay, D.; Shafer, R.W.; Richman, D.D. 2014 Update of the drug resistance mutations in HIV-1. *Top Antivir. Med.* **2014**, *22*, 642–650.
67. Li, J.Z.; Paredes, R.; Ribaud, H.J.; Svarovskaia, E.S.; Metzner, K.J.; Kozal, M.J.; Hullsiek, K.H.; Balduin, M.; Jakobsen, M.R.; Geretti, A.M.; *et al.* Low-frequency HIV-1 drug resistance mutations and risk of NNRTI-based antiretroviral treatment failure: A systematic review and pooled analysis. *JAMA* **2011**, *305*, 1327–1335.
68. Gega, A.; Kozal, M.J. New technology to detect low-level drug-resistant HIV variants. *Future Virol.* **2011**, *6*, 17–26.
69. Gianella, S.; Richman, D.D. Minority variants of drug-resistant HIV. *J. Infect. Dis.* **2010**, *202*, 657–666.
70. Codoñer, F.M.; Pou, C.; Thielen, A.; García, F.; Delgado, R.; Dalmau, D.; Álvarez-Tejado, M.; Ruiz, L.; Clotet, B.; Paredes, R. Added value of deep sequencing relative to population sequencing in heavily pre-treated HIV-1-infected subjects. *PLoS ONE* **2011**, *6*, doi:10.1371/journal.pone.0019461.
71. Mohamed, S.; Penaranda, G.; Gonzalez, D.; Camus, C.; Khiri, H.; Boulmé, R.; Sayada, C.; Philibert, P.; Olive, D.; Halfon, P. Comparison of ultra-deep versus Sanger sequencing detection of minority mutations on the HIV-1 drug resistance interpretations after virological failure. *AIDS* **2014**, *28*, 1315–1324.



72. Todesco, E.; Rodriguez, C.; Morand-Joubert, L.; Mercier-Darty, M.; Desire, N.; Wirten, M.; Girard, P.M.; Katlama, C.; Calvez, V.; Marcelin, A.G. Improved detection of resistance at failure to a tenofovir, emtricitabine and efavirenz regimen by ultradeep sequencing. *J. Antimicrob. Chemother.* **2015**, *70*, 1503–1506.
73. Pou, C.; Noguera-Julian, M.; Pérez-Álvarez, S.; García, F.; Delgado, R.; Dalmau, D.; Álvarez-Tejado, M.; Gonzalez, D.; Sayada, C.; Chueca, N.; *et al.* Improved prediction of salvage antiretroviral therapy outcomes using ultrasensitive HIV-1 drug resistance testing. *Clin. Infect. Dis.* **2014**, *59*, 578–588.
74. Fun, A.; Wensing, A.M.J.; Verheyen, J.; Nijhuis, M. Human immunodeficiency virus gag and protease: Partners in resistance. *Retrovirology* **2012**, *9*, doi:10.1186/1742-4690-9-63.
75. Flynn, W.F.; Chang, M.W.; Tan, Z.; Oliveira, G.; Yuan, J.; Okulicz, J.F.; Torbett, B.E.; Levy, R.M. Deep sequencing of protease inhibitor resistant HIV patient isolates reveals patterns of correlated mutations in Gag and protease. *PLoS Comput. Biol.* **2015**, *11*, doi:10.1371/journal.pcbi.1004249.
76. Zagordi, O.; Däumer, M.; Beisel, C.; Beerenwinkel, N. Read length versus depth of coverage for viral quasispecies reconstruction. *PLoS ONE* **2012**, *7*, doi:10.1371/journal.pone.0047046.
77. Giallonardo, F.D.; Töpfer, A.; Rey, M.; Prabhakaran, S.; Duport, Y.; Leemann, C.; Schmutz, S.; Campbell, N.K.; Joos, B.; Lecca, M.R.; *et al.* Full-length haplotype reconstruction to infer the structure of heterogeneous virus populations. *Nucleic Acids Res.* **2014**, *42*, doi:10.1093/nar/gku537.
78. Shao, W.; Boltz, V.F.; Spindler, J.E.; Kearney, M.F.; Maldarelli, F.; Mellors, J.W.; Stewart, C.; Volfovsky, N.; Levitsky, A.; Stephens, R.M.; *et al.* Analysis of 454 sequencing error rate, error sources, and artifact recombination for detection of Low-frequency drug resistance mutations in HIV-1 DNA. *Retrovirology* **2013**, *10*, doi: 10.1186/1742-4690-10-18.
79. Mild, M.; Hedskog, C.; Jernberg, J.; Albert, J. Performance of ultra-deep pyrosequencing in analysis of HIV-1 pol gene variation. *PLoS ONE* **2011**, *6*, doi:10.1371/journal.pone.0022741.



© 2016 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons by Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).